Journal of Multivariate Analysis **79**, 114–137 (2001)
doi:10.1006/jmva.2000.1958, available online at http://www.idealibrary.com on IDEAL®

Estimation of the Asymptotic Variance of Kernel Density Estimators for Continuous Time Processes

Armelle Guillou and Florence Merlevède

Université Paris VI, Paris, France
E-mail: guillou@ccr.jussieu.fr; merleve@ccr.jussieu.fr

Received April 26, 1999; published online June 19, 2001

In order to construct confidence sets for a marginal density f of a strictly station-

In this paper we address the question of nonparametric estimation of the asymptotic variance of $\sqrt{T}\,\hat{f}_T$, an unknown quantity dependent on f. We construct two estimators and study their asymptotic properties. © 2001 Academic Press

AMS 1991 subject classifications: 62G05; 60G10.

Key words and phrases: kernel estimator; continuous processes; strong mixing sequences; confidence sets.

1. INTRODUCTION

Recently, some surprising parametric rates that appear in the nonparametric estimation of probability densities for continuous time processes came to the attention of the mathematical community. This is possible if local irregularities of sample paths provide some additional information to the statistician.

We refer to Castellana and Leadbetter (1986), Bosq (1993, 1995, 1998), Cheze-Payaud (1994), Kutoyants (1997), and Blanke (1996), among others, for results of this kind and various examples such as the Ornstein–Uhlenbeck process and solutions of some stochastic differential equations.

Throughout this paper, $X=(X_t, t\in\mathbb{R})$ will be a $\mathbb{R}$-valued continuous time process defined on a probability space $(\Omega, \mathscr{A}, P)$. We assume that X is a measurable strictly stationary process with an unknown marginal density f.

We wish to estimate f from the data $(X_t, 0\leqslant t\leqslant T)$.

0047-259X/01 $35.00

In the sequel, we will call a *kernel*, an application K_d, from $\mathbb{R}^d$ to $\mathbb{R}$, which is a bounded, continuous symmetric density with respect to Lebesgue measure and such that

$$\lim_{\|u\| \to \infty} \|u\| \, K_d(u) = 0 \qquad \text{and} \qquad \int_{\mathbb{R}^d} \|u\|^2 \, K_d(u) \, du < +\infty,$$

where $\|\cdot\|$ denotes the usual norm in $\mathbb{R}^d$.

The kernel density estimator is defined as

$$\hat{f}_T(x) = \frac{1}{Th_T} \int_0^T K_1\left(\frac{x - X_t}{h_T}\right) dt, \qquad x \in \mathbb{R}, \tag{1.1}$$

where $h_T \to 0(+)$ as $T \to \infty$ and K_1 is a kernel over $\mathbb{R}$.

The following notation and assumptions will be used throughout the paper and denoted in all the sequel by conditions (A_0). The joint density of X_0 and X_u will be denoted by f_{X_0, X_u} and assumed to exist for all $u \neq 0$. It will be assumed that $g_u(x, y) := f_{X_0, X_u}(x, y) - f(x) f(y)$ for all x, y, $u \mapsto \|g_u\|_\infty \in L_1((0, \infty))$ and that $g_u(\cdot, \cdot)$ will be continuous at (x, x) for each $u > 0$. Let us now recall the well-known Castellana and Leadbetter's theorem (1986).

THEOREM 1.1 (Castellana and Leadbetter (1986)). *Under* (A_0), *we have, as* $T \to \infty$,

$$T \operatorname{Var}(\hat{f}_T(x)) \to G(x), \qquad x \in \mathbb{R}, \tag{1.2}$$

where $G(x) := 2 \int_0^{+\infty} g_u(x, x) \, du$.

The result means that if the distribution of (X_0, X_u) is not too close to a singular distribution for $|u|$ small, then $\hat{f}_T$ converges at the "full rate": $\frac{1}{T}$.

In this paper we address the question of the nonparametric estimation of the asymptotic variance of $\sqrt{T} \hat{f}_T(x)$, as $T \to \infty$. Having such a variance estimate is required in order to get confidence sets for $f(x)$ via the Central Limit Theorem. Indeed, if one wants to construct confidence sets for a density, a first step is to find sufficient conditions under which

$$\sqrt{T} \, (\hat{f}_T(x) - f(x)) \xrightarrow{\mathscr{D}} \mathscr{N}(0, G(x)), \qquad \text{as} \quad T \to \infty. \tag{1.3}$$

Concerning this problem, Bosq *et al.* (1999) have shown that (1.3) is true under some mild mixing conditions (related results under strong mixing conditions may be found in Castellana (1982)). Their main assumption is that there exists $1 < a \leqslant \infty$ such that the strong mixing coefficients of $(X_t, t \in \mathbb{R})$ satisfy the following rate of convergence: $\sum_{k \geqslant 1} k \alpha_k^{(a-1)/a} < \infty$. Obviously, the worst case occurs when a is close to 1. But, in this context, the rate of convergence will be nevertheless satisfied if the process is supposed to be geometrically strong mixing. It is interesting to indicate that, recently, Kutoyants (1997) has obtained a Central Limit Theorem concerning density estimation for diffusion processes and that these processes are precisely geometrically strong mixing.

However, the asymptotic normal distribution in (1.3) cannot be used to yield approximate practical confidence sets since the integrand $\int_0^{+\infty} g_u(x, x)\, du$ is unknown. As far as we know, no estimator of this quantity has yet been studied. Therefore the asymptotic variance must be estimated in order for the Central Limit Theorem to be used. To this effect we introduce two estimators: one simply by replacing $g_u(x, x)$ in the integrand by a kernel-type estimator; the other is a kind of empirical variance.

The remainder of our paper is organized as follows. After presenting our basic notations, we proceed in Section 2 to formally define two variance estimators and show that they are asymptotically unbiased. In Section 3, we study the bias in the Gaussian case and we provide it in a more general context. Finally, in Section 4, we exhibit the mean squared error; all the detailed proofs of our results are postponed to Section 5.

2. CONSTRUCTION OF ESTIMATORS

We shall now describe a way to obtain an estimator of $G(x)$ by plugging the unknown quantities in the integrand by their kernel-type estimators. Thus, the first estimator of $g_u(x, x)$, namely $\hat{g}_{u,T}(x, x)$, is given by

$$\hat{g}_{u,T}(x, x) := \hat{f}_{u,T}(x, x) - \hat{f}_T^2(x), \qquad x \in \mathbb{R},$$

where $\hat{f}_T(x)$ is defined by (1.1) and $\hat{f}_{u,T}(y, z)$ as

$$\hat{f}_{u,T}(y, z) = \frac{1}{T-u} \int_0^{T-u} \frac{1}{h_T^2} K_2\left(\frac{y - X_t}{h_T}, \frac{z - X_{t+u}}{h_T}\right) dt,$$

$$(y, z) \in \mathbb{R}^2, \qquad 0 \leqslant u < T,$$

where $h_T \to 0(+)$ as $T \to \infty$ and K_2 is a kernel defined over $\mathbb{R}^2$.

Then we can define an estimator of $G(x)$; namely, $\hat{G}_T(x)$, as

$$\hat{G}_T(x) = 2 \int_{\varepsilon_T}^{b_T} \hat{g}_{u,T}(x, x)\, du, \tag{2.1}$$

where $\varepsilon_T \to 0$ and $b_T \to \infty$ as $T \to \infty$.

On the other hand, Theorem 1.1 itself seems to suggest another way to look at this problem of estimation. Indeed, let us introduce

(2.2)
$$Y_{n,j}(x) = \frac{1}{\delta}\left\{\int_{(j-1)\delta}^{j\delta} \frac{1}{h_T} K_1\left(\frac{x - X_t}{h_T}\right) dt - \mathbb{E}\int_{(j-1)\delta}^{j\delta} \frac{1}{h_T} K_1\left(\frac{x - X_t}{h_T}\right) dt\right\},$$
$$j \in \mathbb{Z},$$

where $n\delta = T$, $n = [T]$ $(T \geqslant 1)$ and, consequently, $2 > \delta \geqslant 1$. Throughout the paper, square brackets designate the integer part, as usual.

Since $\bar{Y}_n(x) := \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}(x) = \hat{f}_T(x) - \mathbb{E}\hat{f}_T(x)$, under (A_0), as $T \to \infty$,

$$T \operatorname{Var} \bar{Y}_n(x) \to G(x), \qquad x \in \mathbb{R}.$$

Thus, there is a natural way to estimate $G(x)$ from $Y_{n,1}(x), ..., Y_{n,n}(x)$; namely, to look at the sample variability of $\sqrt{\delta/b_T}\sum_{k=i}^{i+b_T-1} Y_{n,k}(x)$, for $i = 1, ..., n - b_T + 1$. This idea leads to the natural estimate

$$\widehat{V_{b_T}}(x) = \frac{1}{n - b_T + 1} \sum_{i=1}^{n-b_T+1} \left(\sqrt{\frac{\delta}{b_T}} \sum_{k=i}^{i+b_T-1} Y_{n,k}(x)\right)^2, \tag{2.3}$$

which can be viewed as an empirical variance. The proposed estimator depends on a design integer parameter b_T that tends to infinity as the sample size T increases, with $b_T < T$.

Note that the estimator defined in (2.3) is inspired by papers from Carlstein (1986) and Politis and Romano (1993). Indeed, in the paper by Carlstein (1986), a quite similar variance estimator was introduced for a general statistic based on subseries values and consistency as well as asymptotic normality were proved in the case of estimating a parameter of a finite-dimensional marginal of the stationary process. By taking advantage of the special structure associated with a general statistic, Politis and Romano (1993) were able to further obtain results on the bias and the variance of the variance estimator as well. However, the central contribution of their paper was to allow for the possibility of working with an estimator consistent for a parameter of the whole (infinite-dimensional) distribution of the process.

Let us introduce now some notations which will be used to show that our two estimators are asymptotically unbiased and to establish their mean squared error.

In some situations, we will consider the space $\mathbb{H}_{k,\lambda}$ of the *kernels of order* (k, λ) in $\mathbb{R}$, $(k \in \mathbb{N}^*, 0 < \lambda \leqslant 1)$; that is, the space of mappings $K_1 \colon \mathbb{R} \to \mathbb{R}$, bounded, integrable, such that $\int_{\mathbb{R}} K_1(v)\, dv = 1$ and satisfying the conditions

$$
\left\{
\begin{array}{l}
\int_{\mathbb{R}} |s|^{k+\lambda} \, |K_1(s)| \, ds < \infty \\
\qquad\qquad \text{and} \\
\int_{\mathbb{R}} s^i K_1(s) \, ds = 0, \qquad 1 \leqslant i \leqslant k.
\end{array}
\right. \tag{2.4}
$$

If $k=0$, (2.4) is simply replaced by

$$
\int_{\mathbb{R}} |s|^{\lambda} \, |K_1(s)| \, ds < \infty. \tag{2.5}
$$

Note that a kernel is not only a positive kernel of order (1, 1) but also a positive kernel of order (0, 1).

Concerning the properties of f, $f_u := f_{X_0, X_u}$ and g_u, we introduce the space $C_r^d(l)$ ($r = k + \lambda$, $k \in \mathbb{N}$, $0 < \lambda \leqslant 1$) of real valued functions h, defined on $\mathbb{R}^d$, which are k times differentiable and such that

$$
\left| \frac{\partial^k h}{\partial x_i^k}(x') - \frac{\partial^k h}{\partial x_i^k}(x) \right| \leqslant l \, \|x' - x\|^{\lambda}, \tag{2.6}
$$

$x, x' \in \mathbb{R}^d$ and $i = 1, ..., d$.

Let us first introduce assumptions which will be used to show that our first estimator of the asymptotic variance defined by (2.1) and called, in what follows, the "plug-in" estimator, is asymptotically unbiased.

(A_1) • $f \in C_1^1(l)$.

(A_2) • $f_u \in C_1^2(l(u))$ and $\int_{\varepsilon_T}^{b_T} l(u) \, du = O(b_T^{\alpha}) + O(\varepsilon_T^{\beta})$; $(\alpha, \beta) \in \mathbb{R}^2$.

The following theorems give conditions under which our estimators are asymptotically unbiased.

THEOREM 2.1. *Suppose* (A_0), (A_1), (A_2) *are satisfied,* $h_T = O(\frac{1}{T})$ *and* $\varepsilon_T^{\beta}/T \to 0$. *If either* $\alpha > 0$ *and* $b_T = o(\min(T, T^{1/\alpha}))$ *or* $\alpha \leqslant 0$ *and* $b_T = o(T)$, *then we have*

$$
\mathbb{E}\hat{G}_T(x) \to G(x), \qquad \text{as} \quad T \to \infty.
$$

THEOREM 2.2. *If* (A_0) *holds, we have*

$$
\mathbb{E}\widehat{V_{b_T}}(x) \to G(x), \qquad \text{as} \quad T \to \infty.
$$

The "empirical" estimator is shown to be asymptotically unbiased under mild conditions; this is not the case for the "plug-in" one.

In the special case where X is a Gaussian process, it is possible to exhibit convergence rates for the bias of our estimators. This is the aim of the next section where we also provide the bias of the "empirical" estimator in a more general context.

3. STUDY OF THE BIAS

3.1. *The Gaussian Case*

In this section $X=(X_t,\, t\in\mathbb{R})$ will be a Gaussian real centered stationary process, continuous in quadratic mean with variance $\sigma^2>0$ and autocorrelation function $\rho(u)$ which satisfies

$$(3.1)\quad \begin{cases} \bullet\ 0\leqslant\rho(u)<1, & u\neq 0,\\ \bullet\ \rho(u)\leqslant u^{-\theta}, & \theta>1, \quad \text{as}\quad u\to\infty,\\ \bullet\ \rho(u)=1-b\,|u|^{\gamma}+o(u^{\gamma}), & b>0, \quad 0<\gamma<2, \quad \text{as}\quad u\to 0.\end{cases}$$

In this case, it is easy to verify (see Castellana and Leadbetter, 1986) that for all $(y,z)\in\mathbb{R}^2$,

$$(3.2)\qquad |g_u(y,z)|\leqslant c\rho(u)\,\mathbb{I}_{|u|>d}+(e+f\,|u|^{-\gamma/2})\,\mathbb{I}_{|u|\leqslant d,\,u\neq 0},$$

where c, d, e, f are suitable constants.

In what follows, the notation $c_T\simeq d_T$ means that c_T and d_T are of the same asymptotic order, that is, $c_T=O(d_T)$ and $d_T=O(c_T)$.

THEOREM 3.1. (1) *If $\gamma\leqslant 1$, the choice of $\varepsilon_T\simeq b_T^{2(1-\theta)/(2-\gamma)}$, $b_T\simeq T^{1/\theta}$ and $h_T\simeq\frac{1}{T}$ leads to the following asymptotic order of the bias*

$$\mathbb{E}\hat{G}_T(x)-G(x)=O\left(\frac{1}{T^{(1-1/\theta)}}\right).$$

(2) *If $\gamma>1$, the choice of $\varepsilon_T\simeq h_T^{2/\gamma}$, $h_T\simeq(1/b_T)^{\gamma/2(\gamma-1)}$ and $b_T\simeq T^{1/\zeta}$ – where $\zeta=\min(\gamma/2(\gamma-1),\theta)$ – leads to the following asymptotic order of the bias*

$$\mathbb{E}\hat{G}_T(x)-G(x)=O\left(\frac{1}{T^{(1-1/\zeta)}}\right).$$

EXAMPLE 3.1. According to the proof of the above theorem, in the particular case of the Ornstein–Uhlenbeck process, we can establish that the asymptotic order of the bias is $O(\frac{\log T}{T})$.

THEOREM 3.2. *Set* $\eta = \min(1, \theta - 1)$. *If either* $\gamma \leqslant 1$ *and* $h_T \simeq 1/b_T^\eta$ *or* $\gamma > 1$ *and* $h_T \simeq (1/b_T)^{\gamma\eta/(2-\gamma)}$, *then we have*

$$\mathbb{E}\widehat{V_{b_T}}(x) - G(x) = O\left(\frac{1}{b_T^\eta}\right).$$

EXAMPLE 3.2. (1) Notice that in the particular case of the Ornstein–Uhlenbeck process, the sample variance estimator leads to a better asymptotic order of the bias, namely $O(\frac{\log\log T}{T})$, than the "plug-in" one.

(2) Clearly, it is possible to obtain a smaller bound for the bias by choosing the estimator $\widehat{V_{b_T}}(x)$ instead of $\hat{G}_T(x)$. Indeed, we can obtain for all $\theta \geqslant 2$ a bias of order $O(\frac{\log\log T}{T})$, since $b_T < T$.

3.2. *The General Case*

Here we state the bias of the sample variance estimator only in the general case, that is without assuming the process to be Gaussian.

Before stating the result, let us introduce some notations and assumptions.

First, we will suppose that g_u satisfies the following assumptions:

- (H_1) $|g_u(z') - g_u(z)| \leqslant l(u)\,\|z' - z\|$; $z, z' \in \mathbb{R}^2$,
- (H_2) $K_1 \in \mathbb{H}_{k,\lambda}$, $g_u \in C^2_{k+\lambda}(m(u))$, $r = k + \lambda$,
- (H_3) $\int_{\varepsilon_T}^{\delta b_T} m(u)\,du = O(b_T^\rho) + O(\varepsilon_T^\varrho)$; $1 \leqslant \delta < 2$, $(\rho, \varrho) \in \mathbb{R}^2$,

where $l(u)$ and $m(u)$ are functions depending only on u.

On the other hand, X is assumed to be a weakly dependent real-valued continuous time process. The degree of dependence is quantified by the various mixing coefficients (cf. Roussas and Ioannides, 1987). We will particularly make use of Rosenblatt's α-mixing or strong mixing coefficient which is defined for each $u \in \mathbb{R}_+$ as

$$\alpha_0 = 1/4 \qquad \text{and} \qquad \alpha_u = \sup_{A,B} |P(A \cap B) - P(A)\,P(B)| \qquad \text{for} \quad u > 0,$$

where $A \in \sigma(X_s, s \leqslant 0)$ and $B \in \sigma(X_s, s \geqslant u)$.

The (strictly stationary) real-valued continuous time process X is said to be "strong mixing" if $\alpha_u \to 0$ as $u \to \infty$.

Finally we will suppose that

- (H_4) $\int_{\delta b_T}^{\infty} (1 + l(u))\,\alpha_u^{1/3}\,du = O(b_T^\mu)$; $1 \leqslant \delta < 2$, $\mu \in \mathbb{R}_-^*$,
- (H_5) $\int_0^{\varepsilon_T} l(u)\,du = O(\varepsilon_T^\nu)$; $\nu \in \mathbb{R}_+^*$,
- (H_6) $\int_{\varepsilon_T}^{\delta b_T} \frac{u}{b_T}(1 + l(u))\,\alpha_u^{1/3}\,du = O(b_T^\vartheta) + O(\varepsilon_T^\psi)$; $1 \leqslant \delta < 2$, $(\vartheta, \psi) \in \overline{\mathbb{R}}_-^* \times \overline{\mathbb{R}}_+^*$.

Now we state the result:

THEOREM 3.3. *Assume that* (A_0) *and* (H_1)–(H_6) *hold. Then we have*

$$\mathbb{E}\widehat{V_{b_T}}(x) - 2\int_0^{+\infty} g_u(x, x)\, du = O(h_T^r b_T^\rho) + O(h_T^r \varepsilon_T^\varrho) + O(b_T^\xi) + O(\varepsilon_T^\tau),$$

where $\xi = \max(\mu, \vartheta)$ *and* $\tau = \min(1, v, \psi)$.

Now let us see what these assumptions mean in the Gaussian case. We assume that X is a real Gaussian process as defined in Subsection 3.1. In this case, according to Rozanov (1967, p. 181), the strong mixing coefficient α_u and the autocorrelation function $\rho(u)$ are equivalent; namely,

$$\alpha_u \leqslant \rho(u) \leqslant 2\pi\alpha_u. \tag{3.3}$$

Moreover, by using (5.12), we have

$$l(u) = (K_1 + K_2\, |u|^{-\gamma})\, \mathbb{1}_{|u| < \varepsilon,\, u \neq 0} + K_3 \rho(u)\, \mathbb{1}_{|u| \geqslant \varepsilon}$$

and $g_u \in C^2_{k+1}(m(u))$ with

$$m(u) = (K'_1 + K'_2\, |u|^{-\gamma(k/2+1)})\, \mathbb{1}_{|u| < \varepsilon,\, u \neq 0} + K'_3 \rho(u)\, \mathbb{1}_{|u| \geqslant \varepsilon},$$

where K_i, K'_i, $i = 1, 2, 3$, and ε are suitable constants.

These considerations lead to

$$\int_{\varepsilon_T}^{\delta b_T} m(u)\, du = O(1) + O(\varepsilon_T^{1-\gamma(k/2+1)}),$$

$$\int_{\delta b_T}^{\infty} (1 + l(u))\, \alpha_u^{1/3}\, du = O(b_T^{1-\theta/3}),$$

$$\int_0^{\varepsilon_T} l(u)\, du = O(\varepsilon_T^{1-\gamma})$$

and

$$\int_{\varepsilon_T}^{\delta b_T} \frac{u}{b_T} (1 + l(u))\, \alpha_u^{1/3}\, du = O\left(\max\left(\frac{1}{b_T}, \frac{1}{b_T^{\theta/3-1}}\right)\right).$$

Then in order for the bias to tend to zero, it is necessary that $\theta > 3$ and $\gamma < 1$. These additive conditions are due to the fact that we use inequality (5.18) (in Theorem 3.3) instead of (3.2) (in Theorem 3.2) to bound $\|g_u\|_\infty$. However, (5.18) is important in the sense that it is valid as soon as (H_1) holds, whatever process X is.

4. STUDY OF THE MEAN SQUARED ERROR

Example 3.2 has particularly shown that the sample variance estimator has better properties, in term of bias, than the "plug-in" one. Thus, it seems natural to simply focus on the study of the mean squared error, say MSE, of $\widehat{V_{b_T}}$.

We will first give results about the asymptotic order of the variance, according to the classical decomposition of the MSE between the square of the bias and the variance. Furthermore, let us notice now that the "plug-in" estimator is inconvenient to work with, not only in term of bias, but also in term of variance since study of such a variance would require assumptions on the joint density of $(X_t, X_{t+j}, X_s, X_{s+k})$.

Bearing in mind that our objective is to construct confidence sets for the density by using the Central Limit Theorem, it seems natural to suppose that there exists $1 < a \leqslant \infty$, such that the strong mixing coefficients of $(X_t, t \in \mathbb{R})$ satisfy the following rate of convergence: $\sum_{k \geqslant 1} k\alpha_k^{(a-1)/a} < \infty$ since this assumption is precisely the one used by Bosq *et al.* (1999) to establish (1.3).

THEOREM 4.1. *Assume that* (A_0) *holds and that there exists* $1 < a < \infty$ *such that*

$$\sum_{k \geqslant 1} k\alpha_k^{(a-1)/a} < \infty. \tag{4.1}$$

Then we have

$$\operatorname{Var}(\widehat{V_{b_T}}(x)) = O\left(\frac{b_T}{T}\left(\frac{1}{h_T}\right)^{(4a-2)/a}\right) + O\left(\frac{b_T}{T}\frac{1}{b_T^{1/(a-1)}}\left(\frac{1}{h_T}\right)^{(4a-1)/a}\right). \tag{4.2}$$

If (4.1) *is replaced by*

$$\sum_{k \geqslant 1} k\alpha_k < \infty, \tag{4.3}$$

then we have

$$\operatorname{Var}(\widehat{V_{b_T}}(x)) = O\left(\log T \frac{b_T}{T}\frac{1}{h_T^4}\right). \tag{4.4}$$

From Theorem 2.2 and Theorem 4.1, we can easily deduce the following corollary:

COROLLARY 4.1. *Assume that* (A_0) *and* (4.1) *hold. In addition suppose that*

$$b_T = o(Th_T^{(4a-2)/a}) \tag{4.5}$$

and that

(4.6) $\quad b_T \geqslant C_1 h_T^{-(a-1)/a}, \quad$ *where* C_1 *is a strictly positive constant.*

Then we have

$$\lim_{T\to\infty} \mathbb{E}(\widehat{V_{b_T}}(x) - G(x))^2 = 0. \tag{4.7}$$

Now if (4.1) *is replaced by* (4.3) *and* (4.5) *and* (4.6) *by*

$$b_T = o\left(T\frac{h_T^4}{\log T}\right), \tag{4.8}$$

then (4.7) *holds.*

As a by-product, $\widehat{V_{b_T}}(x)$ is consistent in probability; then, by using Theorem 2.2 in Bosq *et al.* (1999), we have the following Central Limit Theorem:

COROLLARY 4.2. *Assume that* (A_0), (4.1), (4.6) *and the following conditions are satisfied*:

$$h_T \geqslant C_2 T^{-a^2/(3a-1)(2a-1)}, \tag{4.9}$$

where C_2 *is a strictly positive constant,*

$$b_T = o(T^{(a-1)/(3a-1)}). \tag{4.10}$$

In addition suppose that $f \in C_2^1(l)$, $h_T = o(T^{-1/4})$ *and that* $G(x) > 0$. *Then*

$$\sqrt{T}\,\frac{\widehat{f_T}(x) - f(x)}{\sqrt{\widehat{V_{b_T}}(x)}} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1), \qquad \text{as} \quad T\to\infty, \quad x\in\mathbb{R}. \tag{4.11}$$

Remark 4.1. (1) Note that it is easy to show that conditions (4.6), (4.9), and (4.10) are compatible since $a > 1$. Moreover let us notice that if $h_T = \frac{c}{\log\log T} T^{-1/4}$, $c > 0$, then the choice $h_T = o(T^{-1/4})$ is compatible with (4.9), if (4.1) is satisfied for $a \in]1, (5+\sqrt{17})/4[$.

(2) Sufficient conditions insuring that $G(x) > 0$ have been exhibited in Lemma 3.1 in Blanke and Merlevède (2000).

Now we state results about the mean squared error in the special case of Gaussian processes. Note that even if the process is not Gaussian but strong mixing, it is possible to exhibit a mean squared error by using Theorem 3.3 combined with Theorem 4.1.

COROLLARY 4.3. *Assume that X is a Gaussian real stationary process as defined in Subsection* 3.1 *and suppose that there exists $1 < a < \infty$ such that $\theta > \frac{2a}{a-1}$. Set $\alpha = \frac{2a-1}{a}$.*

(1) *If $\gamma \leqslant 1$ we set $k = [\frac{2(1-\gamma)}{\gamma}]$. Then assuming that $K_1 \in \mathbb{H}_{k,1}$, the choice of $h_T \simeq (1/b_T)^{1/(k+1)}$ and $b_T \simeq T^{\kappa_1}$ minimizes the asymptotic mean squared error, yielding*

$$MSE := \mathbb{E}(\widehat{V_{b_T}}(x) - G(x))^2 = O\left(\left(\frac{1}{T}\right)^{2\kappa_1}\right),$$

where $\kappa_1 := \frac{k+1}{3k+2\alpha+3}$.

(2) *If $\gamma > 1$ and if $\alpha \leqslant \frac{2}{\gamma}$, the choice of $h_T \simeq (1/b_T)^{\gamma/(2-\gamma)}$ and $b_T \simeq T^{\kappa_2}$ minimizes the asymptotic mean squared error, yielding*

$$MSE = O\left(\left(\frac{1}{T}\right)^{2\kappa_2}\right),$$

where $\kappa_2 := \frac{2-\gamma}{3(2-\gamma)+2\alpha\gamma}$.

(3) *If $\gamma > 1$ and $\alpha > \frac{2}{\gamma}$, the only difference is that we choose $b_T \simeq T^{\kappa_3}$, which leads to*

$$MSE = O\left(\left(\frac{1}{T}\right)^{2\kappa_3}\right),$$

where $\kappa_3 := (2-\gamma)(\alpha-1)/((2-\gamma)(4\alpha-5) + \gamma(\alpha+2)(\alpha-1))$.

Remark 4.2. (1) In the particular case of the Ornstein–Uhlenbeck process, (4.1) is satisfied for a close to 1. We can establish an asymptotic mean squared error close to $O(1/T^{2/5})$.

(2) It is clear that the asymptotic order of the mean squared error is decreasing with k. Then for a Gaussian process satisfying (3.1) with γ close to zero, k can be chosen arbitrarily large, yielding an asymptotic order of the mean squared error close to $O(1/T^{2/3})$, as soon as there exists $1 < a < \infty$ such that $\theta > \frac{2a}{a-1}$.

Remark 4.3. Take the case where we only know that $\theta > 2$ and $\gamma \leqslant 1$. If $K_1 \in \mathbb{H}_{k,1}$—where $k = [\frac{2(1-\gamma)}{\gamma}]$—then we can establish that the asymptotic order of the mean squared error is $O((\frac{\log T}{T})^{2(k+1)/(3k+7)})$, which is close to $O((\frac{\log T}{T})^{2/3})$ as soon as γ is closely related to zero.

Let us give some remarks concerning the magnitude of the mean squared error given in Corollary 4.3. Recently Blanke and Merlevède (2000) have shown that we can construct an estimator of $G(x)$ which converges in quadratic mean at the rate $T^{-2p/(2p+1)}$, where $p \geqslant 1$ depends on the process X. More precisely, they use the local time, $l_T(x)$, to derive an unbiased estimator of the unknown density $f(x)$—see Bosq (1998, Chap. 6) for more details—and then, taking into account its additivity property, they exhibit an estimator, $\hat{G}_T^*(x)$, of $G(x)$. The aforementioned rate is obtained under the assumption that $\mathbb{E}l_T^4(x)$ is finite and conditions which represent combinations between the mixing rate of the process X and a "tail" assumption on the marginal distribution of the local time. They also exhibit intermediate rates under a weaker assumption; namely, $\mathbb{E}l_T^{2+\delta}(x) < \infty$ with $0 \leqslant \delta < 2$. According to Lemma 6.3 in Bosq (1998), for a given Gaussian process X assuming to be in addition geometrically strong mixing (like the Ornstein–Uhlenbeck process), its corresponding local time admits moments of exponential order and then $\hat{G}_T^*(x)$ leads to better results than $\widehat{V_{b_T}}(x)$ in term of quadratic mean. Such a conclusion is fortunatly not the same for every process X. Indeed in general it is difficult to compare the conditions required in this paper with the assumptions made in Blanke and Merlevède (2000). In fact, the choice between $\widehat{V_{b_T}}(x)$ or $\hat{G}_T^*(x)$ as an estimator of $G(x)$ depends strongly on what it is assumed on the process X. However, it is clear that $\widehat{V_{b_T}}(x)$ is more attractive than $\hat{G}_T^*(x)$ for practical reasons, taking into account the fact that it is entirely based on the observations of X.

Now, bearing in mind that a practical confidence set for $f(x)$ is our original problem, one can see that it can be solved by using the Central Limit Theorem given in Corollary 4.2. The confidence region obtained by this proposed method is accurate up to the first-order. In view of this, one may naturally ask whether we can obtain, for example by using resampling techniques, second-order optimality as the independent and identically distributed (i.i.d.) bootstrap of Efron (1979). Indeed, it is well known that, in the i.i.d. set up, bootstrap provides more accurate approximations to the distributions of many regular statistics than the classical large sample approximations. However, inferences methods for i.i.d. data, or, more generally, independent data are simply not consistent when the underlying sequence is dependent. Therefore, the resampling and subsampling methods need to be modified to be applicable in such cases. In the last few years, several authors have tried to extend Efron's idea for applying it to some specific dependent models. A general approach is to apply resampling to the original data sequence by considering blocks of data rather than single data points as in the i.i.d. setup. The motivation is that within each block the dependence structure of the underlying model is preserved and if the block size is allowed to tend to infinity with the sample size, asymptotically correct inference can ensue. A pioneering work was provided by Carlstein (1986) who used a blocking scheme to approximate the variance of a general statistic. His idea was to divide the original sequence in (non-overlapping) blocks of size $b<n$, recompute the statistic of interest on these blocks and use the sample variance of the block statistics, after some suitable normalization. Later, Künsch (1989) and Liu and Singh (1992) independently introduced the moving blocks bootstrap, which besides variance estimation can also be used to estimate the sampling distribution of a statistic so that confidence intervals or regions for unknown parameters can be constructed. As in the independent case, the key step in establishing the second-order correctness of stationary bootstrap procedure is to derive Edgeworth expansions for the original statistic and its corresponding bootstrapped version. In the dependent situation, such a task is usually rather tedious and is outside the scope of the present paper. It leads nevertheless to further research.

5. PROOFS

5.1. *Proof of Theorem* 2.1

We first consider the decomposition

$$\mathbb{E}\hat{G}_T(x)-G(x)=2\mathbb{E}\int_{\varepsilon_T}^{b_T}[\hat{g}_{u,T}(x,x)-g_u(x,x)]\,du \\ -2\int_0^{\varepsilon_T}g_u(x,x)\,du-2\int_{b_T}^{\infty}g_u(x,x)\,du \tag{5.1}$$

and treat each term separately.

Considering the fact that $u\mapsto\|g_u\|_\infty$ is integrable on $]0,\infty[$, $\varepsilon_T\to 0$, and $b_T\to\infty$ as $T\to\infty$, it is obvious that the two last terms in the above expression tend to zero.

Now the first term, I, may be rewritten as

$$I=2\int_{\varepsilon_T}^{b_T}(\mathbb{E}\hat{f}_{u,T}(x,x)-f_u(x,x))\,du-2(b_T-\varepsilon_T)(\mathbb{E}\hat{f}_T^2(x)-f^2(x)) \\ =: 2\Delta_T-2\delta_T.$$

Let us notice that

$$\frac{\delta_T}{b_T-\varepsilon_T}=\mathbb{E}(\hat{f}_T(x)-f(x))^2+2f(x)\,\mathbb{E}(\hat{f}_T(x)-f(x)).$$

First, since $\int_{\mathbb{R}} K_1(s)\, ds = 1$ and $f \in C_1^1(l)$, we have

$$|\mathbb{E}(\hat{f}_T(x) - f(x))| = \left| \int_{\mathbb{R}} K_1(s)(f(x - h_T s) - f(x))\, ds \right|$$

$$\leqslant l h_T \int_{\mathbb{R}} |s|\, K_1(s)\, ds.$$

But $\int_{\mathbb{R}} |s|\, K_1(s)\, ds < \infty$ since K_1 is a kernel; therefore,

$$\mathbb{E}(\hat{f}_T(x) - f(x)) = O(h_T),$$

which leads to

$$\mathbb{E}(\hat{f}_T(x) - f(x))^2 = O(h_T^2) + O\left(\frac{1}{T}\right),$$

combining the facts that $\mathbb{E}(\hat{f}_T(x) - f(x))^2 = (\mathbb{E}\, \hat{f}_T(x) - f(x))^2 + \operatorname{Var} \hat{f}_T(x)$ and that under (A_0), $\operatorname{Var} \hat{f}_T(x) = O(\frac{1}{T})$.

Thus,

$$\delta_T = O(b_T h_T) + O\left(\frac{b_T}{T}\right).$$

Now, using stationarity and the fact that $\int_{\mathbb{R}^2} K_2(s, t)\, ds\, dt = 1$, we find that

$$\begin{aligned} \Delta_T = & \int_{\varepsilon_T}^{b_T} \left\{ \frac{1}{T-u} \int_0^{T-u} \int_{\mathbb{R}^2} \frac{1}{h_T^2} K_2\left(\frac{x-v}{h_T}, \frac{x-w}{h_T} \right) f_{X_t, X_{t+u}}(v, w) \right. \\ & \left. \times\, dv\, dw\, dt - f_u(x, x) \right\} du \\ = & \int_{\varepsilon_T}^{b_T} \int_{\mathbb{R}^2} K_2(s, t) \{ f_u(x - s h_T, x - t h_T) - f_u(x, x) \}\, ds\, dt\, du. \end{aligned}$$

Then since $f_u \in C_1^2(l(u))$, we have

$$|\Delta_T| \leqslant h_T \int_{\mathbb{R}^2} \|(s, t)\|\, K_2(s, t)\, ds\, dt \int_{\varepsilon_T}^{b_T} l(u)\, du.$$

Using the fact that K_2 is a kernel over $\mathbb{R}^2$, we have $\int_{\mathbb{R}^2} \|(s, t)\|\, K_2(s, t)\, ds\, dt < \infty$. Then (A_2) leads to

$$\Delta_T = O(b_T^{\alpha} h_T) + O(\varepsilon_T^{\beta} h_T).$$

Finally,

$$I = O(b_T h_T) + O\left(\frac{b_T}{T}\right) + O(b_T^{\alpha} h_T) + O(\varepsilon_T^{\beta} h_T),$$

which entails the desired result by our choices of h_T, b_T, and ε_T.

5.2. *Proof of Theorem* 2.2

Let us set $Z_{n,i}(x) = \sqrt{\delta/b_T}\sum_{k=i}^{i+b_T-1} Y_{n,k}(x)$, $i = 1, ..., n - b_T + 1$. By using stationarity, we have

$$\begin{aligned} \mathbb{E}\widehat{V_{b_T}}(x) &= \mathbb{E}Z_{n,1}^2(x) \\ &= \frac{1}{\delta b_T} \operatorname{Var}\left(\int_0^{\delta b_T} \frac{1}{h_T} K_1\left(\frac{x - X_t}{h_T}\right) dt\right) \\ &= 2\int_0^{\delta b_T} \left(1 - \frac{u}{\delta b_T}\right) \operatorname{Cov} \\ &\quad \times \left(\frac{1}{h_T} K_1\left(\frac{x - X_0}{h_T}\right), \frac{1}{h_T} K_1\left(\frac{x - X_u}{h_T}\right)\right) du. \end{aligned} \tag{5.2}$$

Now since $(u, y, z) \mapsto (1/h_T^2)\, K_1(\frac{x-y}{h_T})\, K_1(\frac{x-z}{h_T})(1 - \frac{u}{\delta b_T})\, g_u(y, z)\, \mathbb{1}_{0 \leqslant u \leqslant \delta b_T}$ is integrable, Fubini's Theorem entails that

$$\begin{aligned} \mathbb{E}\widehat{V_{b_T}}(x) &= 2\int_{\mathbb{R}^2} \frac{1}{h_T^2} K_1\left(\frac{x - y}{h_T}\right) K_1\left(\frac{x - z}{h_T}\right) \int_0^{\delta b_T} \\ &\quad \times \left(1 - \frac{u}{\delta b_T}\right) g_u(y, z)\, du\, dy\, dz, \end{aligned} \tag{5.3}$$

which converges to $2\int_0^{+\infty} g_u(x, x)\, du$ under the assumption (A_0) (see Bosq, 1998, p. 99, for more details).

5.3. *Proof of Theorem* 3.1

We again consider decomposition (5.1) and first notice that (3.1) combined with (3.2) yields

$$\int_0^{\varepsilon_T} g_u(x, x)\, du = O(\varepsilon_T^{1-\gamma/2}) \qquad \text{and} \qquad \int_{b_T}^{\infty} g_u(x, x)\, du = O(b_T^{1-\theta}). \tag{5.4}$$

On the other hand, it is easy to verify that (A_1) holds and because of (3.2), (A_0) is satisfied. Now to show (A_2), note that the Taylor formula implies

$$
\begin{aligned}
(5.5)\qquad & f_u(x-y, x-z) - f_u(x, x) \\
&= -y\frac{\partial f_u}{\partial x_1}(x-\theta_1 y, x-\theta_1 z) - z\frac{\partial f_u}{\partial x_2}(x-\theta_1 y, x-\theta_1 z),
\end{aligned}
$$

where $0 < \theta_1 < 1$.

It is easy to show that

$$
(5.6)\qquad \left\| \frac{\partial f_u}{\partial x_1} \right\|_\infty \leqslant \frac{C_1}{1-\rho^2(u)},
$$

where C_1 is a constant.

Obviously, a similar bound is valid for $\|\partial f_u/\partial x_2\|_\infty$. These last considerations combined with (5.5) lead to

$$
|f_u(x-y, x-z) - f_u(x, x)| \leqslant (|y| + |z|)\frac{C_2}{1-\rho^2(u)},
$$

where C_2 is a constant.

Moreover since (3.1) holds, $\int_{\varepsilon_T}^{b_T}(1/(1-\rho^2(u)))\,du = O(b_T) + O(\varepsilon_T^{1-\gamma})$. Thus (A_2) is fulfilled with $l(u) = C_2/(1-\rho^2(u))$. Then according to the proof of Theorem 2.1, we derive that

$$
\begin{aligned}
(5.7)\qquad \mathbb{E}\hat{G}_T(x) - G(x) &= O(h_T b_T) + O(h_T \varepsilon_T^{1-\gamma}) \\
&\quad + O\left(\frac{b_T}{T}\right) + O(\varepsilon_T^{1-\gamma/2}) + O(b_T^{1-\theta}).
\end{aligned}
$$

Finally, our choices of b_T, ε_T and h_T complete the proof.

5.4. *Proof of Theorem* 3.2

Notice first that (5.3) can be rewritten as

$$
\mathbb{E}\widehat{V_{b_T}}(x) = 2\int_{\mathbb{R}^2} K_1(s)\,K_1(t)\int_0^{\delta b_T}\left(1 - \frac{u}{\delta b_T}\right) g_u(x - sh_T, x - th_T)\,du\,ds\,dt.
$$

Then we have the decomposition

$$
\begin{aligned}
(5.8)\qquad \mathbb{E}\widehat{V_{b_T}}(x) - G(x) = {} & 2\int_{\mathbb{R}^2} K_1(s)\, K_1(t) \int_0^{\varepsilon_T} \left(1 - \frac{u}{\delta b_T}\right) \\
& \times g_u(x - sh_T, x - th_T)\, du\, ds\, dt \\
& + 2\int_{\mathbb{R}^2} K_1(s)\, K_1(t) \int_{\varepsilon_T}^{\delta b_T} \left(1 - \frac{u}{\delta b_T}\right) \\
& \times (g_u(x - sh_T, x - th_T) - g_u(x, x))\, du\, ds\, dt \\
& - 2\int_{\varepsilon_T}^{\delta b_T} \frac{u}{\delta b_T}\, g_u(x, x)\, du - 2\int_0^{\varepsilon_T} g_u(x, x)\, du \\
& - 2\int_{\delta b_T}^{\infty} g_u(x, x)\, du,
\end{aligned}
$$

where ε_T is a decreasing sequence to zero which will be specified in what follows.

But since $\int_{\mathbb{R}} K_1(s)\, ds = 1$,

$$
\begin{aligned}
& \int_{\mathbb{R}^2} K_1(s)\, K_1(t) \int_{\varepsilon_T}^{\delta b_T} \left(1 - \frac{u}{\delta b_T}\right) (g_u(x - sh_T, x - th_T) - g_u(x, x))\, du\, ds\, dt \\
& \quad = \int_{\mathbb{R}^2} K_1(s)\, K_1(t) \int_{\varepsilon_T}^{\delta b_T} \left(1 - \frac{u}{\delta b_T}\right) \\
& \qquad \times (g_u(x - sh_T, x - th_T) - g_u(x, x - th_T))\, du\, ds\, dt \\
& \qquad + \int_{\mathbb{R}} K_1(t) \int_{\varepsilon_T}^{\delta b_T} \left(1 - \frac{u}{\delta b_T}\right) (g_u(x, x - th_T) - g_u(x, x))\, du\, dt \\
& \quad =: I_1 + I_2.
\end{aligned}
$$

Now for convenience, we will suppose that $K_1 \in \mathbb{H}_{k,1}$ in what follows since this condition will be required in some proofs.

First, by using the Taylor formula and (2.4), we find that

$$
I_1 = \int_{\mathbb{R}^2} K_1(s)\, K_1(t) \frac{(-sh_T)^k}{k!} \int_{\varepsilon_T}^{\delta b_T} \left(1 - \frac{u}{\delta b_T}\right) \frac{\partial^k g_u}{\partial x_1^k} (x - \theta_1 sh_T, x - th_T)\, du\, ds\, dt,
$$

where $0<\theta_1<1$. Similarly,

$$I_2=\int_{\mathbb{R}} K_1(t)\,\frac{(-th_T)^k}{k!}\int_{\varepsilon_T}^{\delta b_T}\left(1-\frac{u}{\delta b_T}\right)\frac{\partial^k g_u}{\partial x_2^k}\,(x, x-\theta_2 th_T)\,du\,dt,$$

where $0<\theta_2<1$.

Now using again (2.4) we find that

$$I_1=\int_{\mathbb{R}^2} K_1(s)\,K_1(t)\,\frac{(-sh_T)^k}{k!}\int_{\varepsilon_T}^{\delta b_T}\left(1-\frac{u}{\delta b_T}\right)$$
$$\times\left(\frac{\partial^k g_u}{\partial x_1^k}\,(x-\theta_1 sh_T, x-th_T)-\frac{\partial^k g_u}{\partial x_1^k}\,(x,x)\right)du\,ds\,dt$$

and

$$I_2=\int_{\mathbb{R}} K_1(t)\,\frac{(-th_T)^k}{k!}\int_{\varepsilon_T}^{\delta b_T}\left(1-\frac{u}{\delta b_T}\right)\left(\frac{\partial^k g_u}{\partial x_2^k}\,(x, x-\theta_2 th_T)-\frac{\partial^k g_u}{\partial x_2^k}\,(x,x)\right)du\,dt.$$

Now since g_u is k times differentiable for all $k\in\mathbb{N}$, the Taylor formula implies

(5.9)
$$\frac{\partial^k g_u}{\partial x_1^k}\,(x-y, x-z)-\frac{\partial^k g_u}{\partial x_1^k}\,(x,x)$$
$$=-y\,\frac{\partial^{k+1} g_u}{\partial x_1^{k+1}}\,(x-\theta_3 y, x-\theta_3 z)$$
$$-z\,\frac{\partial^{k+1} g_u}{\partial x_1^{k}\,\partial x_2}\,(x-\theta_3 y, x-\theta_3 z),$$

where $0<\theta_3<1$.

But by simple induction for all $i\geqslant 0$, it is easy to derive that

(5.10)
$$\frac{\partial^i g_u}{\partial x_1^i}\,(x,y)$$
$$=\begin{cases}\displaystyle\sum_{l=0}^{i/2} M_{i,l}\left\{\frac{(-x+y\rho(u))^{2l}}{\sqrt{a(u)}\,a^{i/2+l}(u)}\exp(-A(\rho(u)))-x^{2l}\exp(-A(0))\right\}\\ \qquad\text{if } i \text{ is even}\\ \displaystyle\sum_{l=0}^{(i-1)/2} M_{i,l}\left\{\frac{(-x+y\rho(u))^{2l+1}}{\sqrt{a(u)}\,a^{(i+1)/2+l}(u)}\exp(-A(\rho(u)))+x^{2l+1}\exp(-A(0))\right\}\\ \text{otherwise,}\end{cases}$$

where $M_{i,l}$ are suitable constants, $a(u) = 1 - \rho^2(u)$ and $A(v) = (1/2(1 - v^2))(x^2 - 2vxy + y^2)$, for all $(x, y) \in \mathbb{R}^2$. Then, for all $i \geqslant 0$,

$$\left\| \frac{\partial^i g_u}{\partial x_1^i} \right\|_\infty \leqslant (C_{1,i} + C_{2,i}\,|u|^{-\gamma((i+1)/2)})\, \mathbb{I}_{|u| < \varepsilon,\, u \neq 0} + C_{3,i}\,\rho(u)\, \mathbb{I}_{|u| \geqslant \varepsilon}, \tag{5.11}$$

where $C_{1,i}$, $C_{2,i}$, $C_{3,i}$, and ε are suitable constants. Classically, the last term on the right-hand side of the above inequality is obtained by applying the Taylor formula to (5.10) (see, e.g., Castellana and Leadbetter, 1986) whereas a direct bound combined with (3.1) leads to the first one. A similar bound can be obtained for $\|\partial^i g_u / \partial x_1^{i-1}\, \partial x_2\|_\infty$.

These last considerations combined with (5.9) show that

$$\begin{aligned} \left| \frac{\partial^k g_u}{\partial x_1^k}(x-y, x-z) - \frac{\partial^k g_u}{\partial x_1^k}(x, x) \right| &\leqslant (|y| + |z|)\{(\tilde{C}_{1,k+1} + \tilde{C}_{2,k+1}\,|u|^{-\gamma(k/2+1)}) \\ &\quad \times \mathbb{I}_{|u| < \varepsilon,\, u \neq 0} + \tilde{C}_{3,k+1}\,\rho(u)\, \mathbb{I}_{|u| \geqslant \varepsilon}\}. \end{aligned} \tag{5.12}$$

A similar bound is valid for $|(\partial^k g_u / \partial x_2^k)(x-y, x-z) - (\partial^k g_u / \partial x_2^k)(x, x)|$.

Thus (5.12) combined with (2.4) and (3.1) leads to

$$I_1 + I_2 = O(h_T^{k+1}) + O(h_T^{k+1}\,\varepsilon_T^{1-\gamma(k/2+1)}). \tag{5.13}$$

On the other hand, taking into account (3.1) and (3.2), we easily obtain

$$\int_{\varepsilon_T}^{\delta b_T} \frac{u}{\delta b_T}\, g_u(x, x)\, du = O\left(\frac{1}{b_T^\eta}\right), \qquad \text{where} \quad \eta := \min(1, \theta - 1) \tag{5.14}$$

and since it is clear that $\int_{\mathbb{R}} |K_1(s)|\, ds < \infty$,

$$\int_{\mathbb{R}^2} K_1(s)\, K_1(t) \int_0^{\varepsilon_T} \left(1 - \frac{u}{\delta b_T}\right) g_u(x - sh_T, x - th_T)\, du\, ds\, dt = O(\varepsilon_T^{1-\gamma/2}). \tag{5.15}$$

Finally, combining (5.13)–(5.15) with (5.4), we derive that

$$\begin{aligned} \mathbb{E}\widehat{V_{b_T}}(x) - G(x) &= O(\varepsilon_T^{1-\gamma/2}) + O\left(\frac{1}{b_T^\eta}\right) + O(h_T^{k+1}) \\ &\quad + O(h_T^{k+1}\,\varepsilon_T^{1-\gamma(k/2+1)}). \end{aligned} \tag{5.16}$$

Now, we return to the case where K_1 is only a usual kernel; then we can take $k = 0$.

First, if $\gamma \leqslant 1$, we choose $\varepsilon_T \simeq (1/b_T)^{2\eta/(2-\gamma)}$ which entails that

$$\mathbb{E}\widehat{V_{b_T}}(x) - G(x) = O\left(\frac{1}{b_T^\eta}\right) + O(h_T)$$

and the choice of $h_T \simeq 1/b_T^\eta$ yields

$$\mathbb{E}\widehat{V_{b_T}}(x) - G(x) = O\left(\frac{1}{b_T^\eta}\right).$$

Now if $\gamma > 1$, we are this time led to choose $\varepsilon_T \simeq h_T^{2/\gamma}$ and $h_T \simeq (1/b_T)^{\gamma\eta/(2-\gamma)}$ to complete the proof.

5.5. *Proof of Theorem* 3.3

We consider decomposition (5.8) and proceed as in the proof of Theorem 3.2. Then (H_2) and (H_3) imply that

$$|I_1 + I_2| = O(b_T^\rho h_T^{k+\lambda}) + O(\varepsilon_T^\varrho h_T^{k+\lambda}). \tag{5.17}$$

Now according to Lemma 1.3 in Bosq (1998), (H_1) entails that

$$\| g_u \|_\infty \leqslant C(1 + l(u))\, \alpha_u^{1/3}, \tag{5.18}$$

where C is a constant; this in turn combined with (H_4), (H_5) and the fact that for all u, $\alpha_u \leqslant \frac{1}{4}$, leads to

$$\int_{\delta b_T}^{\infty} \| g_u \|_\infty \, du = O(b_T^\mu) \qquad \text{and} \qquad \int_0^{\varepsilon_T} \| g_u \|_\infty \, du = O(\varepsilon_T) + O(\varepsilon_T^\nu).$$

Finally the proof is complete by using the fact that under (H_6), $\int_{\varepsilon_T}^{\delta b_T} \frac{u}{b_T} \| g_u \|_\infty \, du = O(b_T^\vartheta) + O(\varepsilon_T^\psi)$.

5.6. *Proof of Theorem* 4.1

In order to prove the theorem, we shall use the following technical lemma which appears in Bosq *et al.* (1999) and which is a consequence of a result of Yokoyama (1980).

LEMMA 5.1. *Suppose that* $\xi := \{\xi_i, i \geqslant 1\}$ *is a strictly stationary strong mixing sequence of zero mean real random variables such that*

$$|\xi_i| \leqslant M \quad \textit{a.s. and} \quad \mathbb{E}\xi_i^2 \leqslant \gamma, \quad \textit{for every} \quad i \geqslant 1, \tag{5.19}$$

where M *and* γ *are constants.*

Assuming that (4.1) *holds, we can find a constant* C *depending only on the strong mixing coefficient of* ξ *such that for every* $n \geqslant 1$

$$\mathbb{E}\left(\sum_{i=1}^{n} \xi_i\right)^4 \leqslant C n^2 M^{(4a-2)/a} \gamma^{1/a}. \tag{5.20}$$

To begin the proof of Theorem 4.1, let us define some quantities closely related to $\widehat{V_{b_T}}(x)$, though easier to work with. So for $l = 1, 2, ..., b_T$, set

$$\widehat{V_{b_T}}^{(l)}(x) := \frac{1}{q_{T,l}} \sum_{i=1}^{q_{T,l}} \left(\sqrt{\frac{\delta}{b_T}} \sum_{k=(i-1)\,b_T+l}^{ib_T+l-1} Y_{n,k}(x)\right)^2,$$

where $q_{T,l} = [(n - b_T - l + 1)/b_T] + 1$.

It is easy to verify that $\widehat{V_{b_T}}(x) = (1/(n - b_T + 1)) \sum_{l=1}^{b_T} q_{T,l} \widehat{V_{b_T}}^{(l)}(x)$. Since all the $q_{T,l}$'s are of the same asymptotic order as $q_T := [(n - b_T)/b_T] + 1 = q_{T,1}$, it is obvious that $\widehat{V_{b_T}}(x) \simeq \frac{1}{b_T} \sum_{l=1}^{b_T} \widehat{V_{b_T}}^{(l)}(x)$.

To show (4.2), we study $\operatorname{Var} \widehat{V_{b_T}}^{(l)}(x)$ for $l = 1, ..., b_T$. In fact, we will only carry out the proof for the case where $l = 1$, the other cases being similar. Let us denote $W_{n,i}(x) = \sqrt{\delta/b_T} \sum_{k=(i-1)\,b_T+1}^{ib_T} Y_{n,k}(x)$.

For all $n \geqslant 1$ and $k \geqslant 1$, let us define the strong mixing coefficient of $Y_n = (Y_{n,i}, i \in \mathbb{Z})$ by

$$\alpha_{n,k}^{(Y_n)} = \sup_{A,B} |P(A \cap B) - P(A)\,P(B)|,$$

where $A \in \sigma(Y_{n,i}, i \leqslant 0)$ and $B \in \sigma(Y_{n,i}, i \geqslant k)$.

This coefficient is uniformly bounded by $\alpha_k^{(Y_n)} := \sup_{n \geqslant 1} \alpha_{n,k}^{(Y_n)}$. Moreover it is clear that for every $k \geqslant 1$, $\alpha_k^{(Y_n)} \leqslant \alpha_{k-1}$. Now since the $W_{n,i}(x)$ are functions of finite blocks of the $Y_{n,k}(x)$, they are obviously $\alpha^{(W_n)}$-mixing with

$$\alpha_k^{(W_n)} \leqslant \alpha_{(k-1)\,b_T} \qquad \text{if} \quad k \geqslant 2. \tag{5.21}$$

Now because of stationarity

$$\begin{aligned} \operatorname{Var}(\widehat{V_{b_T}}^{(1)}(x)) &= \frac{1}{q_T} \operatorname{Var}(W_{n,1}^2(x)) \\ &\quad + \frac{2}{q_T^2} \sum_{i=1}^{q_T-1} (q_T - i) \operatorname{Cov}(W_{n,1}^2(x), W_{n,i+1}^2(x)) \\ &\leqslant \frac{3}{q_T} \operatorname{Var}(W_{n,1}^2(x)) \\ &\quad + \frac{2}{q_T} \sum_{i=2}^{q_T-1} |\operatorname{Cov}(W_{n,1}^2(x), W_{n,i+1}^2(x))|. \end{aligned} \tag{5.22}$$

But the well-known theorem of Ibragimov (see Roussas and Ioannides, 1987) combined with the stationarity gives

$$\mathrm{Cov}(W_{n,1}^2(x), W_{n,i+1}^2(x)) \leqslant 10(\alpha_i^{(W_n)})^{1/2} (\mathbb{E}W_{n,1}^8(x))^{1/2}.$$

Now notice that $\mathbb{E}Y_{n,1}^2(x) = (1/\delta^2)\,\mathrm{Var}((1/h_T)\int_0^\delta K_1((x - X_u/h_T))\,du)$. Then under (A_0),

$$\mathbb{E}Y_{n,1}^2(x) = \frac{2}{\delta^2}\int_{\mathbb{R}^2} \frac{1}{h_T^2} K_1\left(\frac{x-y}{h_T}\right) K_1\left(\frac{x-z}{h_T}\right) \int_0^\delta (\delta - u)\, g_u(y, z)\, du\, dy\, dz,$$

which shows that $\mathbb{E}Y_{n,1}^2(x) < \infty$ for all $n \geqslant 1$. This last consideration, together with the fact that $|Y_{n,i}(x)| \leqslant 2\,\|K_1\|_\infty/h_T$, shows that (5.19) is satisfied. Then we can apply Lemma 5.1 which leads to

$$\begin{aligned}\mathbb{E}W_{n,1}^4(x) &\leqslant \frac{\delta^2}{b_T^2}\,\mathbb{E}\left(\sum_{k=1}^{b_T} Y_{n,k}(x)\right)^4\\ &\leqslant C_1\left(\frac{1}{h_T}\right)^{(4a-2)/a}.\end{aligned}$$

Since $|W_{n,1}(x)| \leqslant 2\sqrt{\delta b_T}\,(\|K_1\|_\infty/h_T)$, using the above result we have on the other hand $\mathbb{E}W_{n,1}^8(x) \leqslant \tilde{C}_1(b_T^2/h_T^4)(1/h_T)^{(4a-2)/a}$.

Finally observe that by (4.1) we get $\alpha_{ib_T} = O([1/(ib_T)]^{2a/(a-1)})$ which implies that $\sum_{i=1}^{q_T} \alpha_{ib_T}^{1/2} = O(1/b_T^{a/(a-1)})$. All these considerations yield

$$\mathrm{Var}(\widehat{V_{b_T}}^{(1)}(x)) = O\left(\frac{b_T}{T}\left(\frac{1}{h_T}\right)^{(4a-2)/a}\right) + O\left(\frac{b_T}{T}\,\frac{1}{b_T^{1/a-1}}\left(\frac{1}{h_T}\right)^{(4a-1)/a}\right),$$

since $q_T \simeq \frac{T}{b_T}$.

Since $\widehat{V_{b_T}}(x) \simeq (1/b_T)\sum_{l=1}^{b_T} \widehat{V_{b_T}}^{(l)}(x)$, we easily complete the proof by the Cauchy–Schwartz inequality.

Now if (4.1) is replaced by (4.3), (4.4) is obtained by using similar arguments as in the proof of (4.2) with a few changes. The powers where a appears should be viewed as the constant obtained by letting $a \to \infty$. The only difference in the proof is that

$$\sum_{i=1}^{q_T} \alpha_{ib_T}^{1/2} = O\left(\sum_{i=1}^{q_T} \frac{1}{ib_T}\right) = O\left(\frac{\log T}{b_T}\right).$$

5.7. *Proof of Corollary* 4.3

Since X is a Gaussian process according to Rozanov (1967, p. 181), $\alpha_k \leqslant \rho(k) \leqslant 2\pi\alpha_k$. Then because of (3.1), (4.1) will be verified if $\theta > \frac{2a}{a-1}$. Thus, obviously, $\eta := \min(1, \theta - 1) = 1$.

Now by using (5.16), (4.2), and the classical decomposition

$$\mathbb{E}(\widehat{V_{b_T}}(x)-G(x))^2=(\mathbb{E}\widehat{V_{b_T}}(x)-G(x))^2+\operatorname{Var}\widehat{V_{b_T}}(x),$$

we obtain in the case where $\gamma \leqslant 1$

$$\mathbb{E}(\widehat{V_{b_T}}(x)-G(x))^2=O(\varepsilon_T^{2-\gamma})+O\left(\frac{1}{b_T^2}\right) + O(h_T^{2(k+1)})+O\left(\frac{b_T}{T}\left(\frac{1}{h_T}\right)^{2\alpha}\right) + O\left(\frac{b_T}{T}\frac{1}{b_T^{(2-\alpha)/(\alpha-1)}}\left(\frac{1}{h_T}\right)^{\alpha+2}\right), \tag{5.23}$$

where $\alpha := \frac{2a-1}{a}$.

Direct computations lead to the desired result. The proofs of cases (2) and (3) are similar. Let us notice that in the two last cases of the corollary, even if we take the kernel in a space $\mathbb{H}_{k,1}$, the mean squared error is unchanged; namely, it does not depend on k.

5.8. *Proof of Remark* 4.3

The only difference here is that we use (4.4) instead of (4.2), which leads to $\varepsilon_T \simeq (1/b_T)^{2/(2-\gamma)}$, $h_T \simeq (1/b_T)^{1/(k+1)}$ and $b_T \simeq (T/\log T)^{(k+1)/(3k+7)}$.

ACKNOWLEDGMENTS

We thank the referees for carefully reading the manuscript and for valuable suggestions which improved the presentation of this paper.

REFERENCES

1. D. Blanke, Estimation de la densité pour des trajectoires non directement observables, *Publ. Inst. Statist. Univ. Paris* **40** (1996), 21–36.
2. D. Blanke and F. Merlevède, Estimation of the asymptotic variance of local time density estimators for continuous time processes, *Math. Methods Statist.* **9** (2000), 270–296.
3. D. Bosq, Vitesses optimales et superoptimales des estimateurs fonctionnels pour les processus à temps continu, *C. R. Acad. Sci. Paris Sér. I Math.* **317** (1993), 1075–1078.
4. D. Bosq, Sur le comportement exotique de l'estimateur à noyau de la densité marginale d'un processus à temps continu, *C. R. Acad. Sci. Paris Sér. I Math.* **320** (1995), 369–372.
5. D. Bosq, "Nonparametric Statistics for Stochastic Processes, Estimation and Prediction," 2nd ed., Lecture Notes in Statistics, Vol. 110, Springer-Verlag, New York/Berlin, 1998.
6. D. Bosq, F. Merlevède, and M. Peligrad, Asymptotic normality for density kernel estimators in discrete and continuous time, *J. Multivariate Anal.* **68** (1999), 78–95.

7. E. Carlstein, The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *Ann. Statist.* **14** (1986), 1171–1179.
8. J. V. Castellana, "Nonparametric Density Estimates for Stationary Stochastic Processes," Ph.D. dissertation, North Carolina University, 1982.
9. J. V. Castellana and M. R. Leadbetter, On smoothed probability density estimation for stationary processes, *Stochastic Process. Appl.* **21** (1986), 179–193.
10. N. Cheze-Payaud, Nonparametric regression and prediction for continuous-time processes, *Publ. Inst. Statist. Univ. Paris* **38** (1994), 37–58.
11. B. Efron, Bootstrap methods: Another look at the jackknife, *Ann. Statist.* **7** (1979), 1–26.
12. H. R. Künsch, The jackknife and the bootstrap for general stationary observations, *Ann. Statist.* **17** (1989), 1217–1241.
13. Y. A. Kutoyants, Some problems of nonparametric estimation by observations of ergodic diffusion process, *Statist. Probab. Lett.* **32** (1997), 311–320.
14. R. Y. Liu and K. Singh, Moving blocks jackknife and bootstrap capture weak dependence, *in* "Exploring the Limits of Bootstrap" (R. LePage and L. Billard, Eds.), pp. 225–248, Wiley, New York, 1992.
15. D. N. Politis and J. P. Romano, On the sample variance of linear statistics derived from mixing sequences, *Stochastic Process. Appl.* **45** (1993), 155–167.
16. G. G. Roussas and D. Ioannides, Moment inequalities for mixing sequences of random variables, *Stochastic Anal. Appl.* **5** (1987), 61–120.
17. Y. A. Rozanov, "Stationary Random Processes," Holden–Day, Oakland, CA, 1967.
18. R. Yokoyama, Moment bounds for stationary mixing sequences, *Z. Wahrsch. Verw. Gebiete* **52** (1980), 45–57.